ARMY RESEARCH LABORATORY

Optimizing Uncertainty in Dempster-Shafer Detectors Fusing Multi-Sensor Data

by Kenneth I. Ranney and Nasser M. Nasrabadi

ARL-TR-4738 **March 2009**

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

Army Research Laboratory
Adelphi, MD 20783-1197

ARL-TR-4738 **March 2009**

Optimizing Uncertainty in Dempster-Shafer Detectors Fusing Multi-Sensor Data

Kenneth I. Ranney and Nasser M. Nasrabadi
Sensors and Electron Devices Directorate, ARL

REPORT DOCUMENTATION PAGE	*Form Approved OMB No. 0704-0188*

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED (From - To)
March 2009	Summary	

4. TITLE AND SUBTITLE

Optimizing Uncertainty in Dempster-Shafer Detectors Fusing Multi-Sensor Data

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Kenneth I. Ranney and Nasser M. Nasrabadi

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

U.S. Army Research Laboratory
ATTN: AMSRD-ARL-SE-RU
2800 Powder Mill Road
Adelphi, MD 20783-1197

8. PERFORMING ORGANIZATION REPORT NUMBER

ARL-TR-4738

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Fusion algorithms based on the Dempster-Shafer (DS) Theory of Evidence lack a universally standard method for automatically assigning probability mass to the "don't know" hypothesis for a particular input. For example, when fusing automatic target detection (ATD) algorithm outputs from multiple sensors, one must associate a measure of uncertainty with the output from the ATD algorithm of each sensor. We describe such a fusion algorithm, developed using the DS formalism, and present a method for automatically determining the required assignment of uncertainty. We also evaluate the entire procedure using simulated data and receiver operating characteristic curves.

15. SUBJECT TERMS

Dempster-Shafer, DS, ATD, algorithm

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE	UU	18	Kenneth I. Ranney
U	U	U			19b. TELEPHONE NUMBER (*Include area code*) (301) 394-0832

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

1. Objective

In this effort, we formulated a technique for automatically determining effective probability mass functions required by Dempster-Shafer (DS) automatic target detection (ATD) algorithms for multi-sensor data fusion. The initial step in this process comprises defining an ATD algorithm performance measure. The second step comprises formulating the automatic probability mass assignments. In short, our goal was to define the distribution of probability mass assigned to the DS "don't know" hypothesis so as to improve the performance of the fused ATD algorithm relative to the results obtained using standard Bayesian approaches.

2. Approach

Since DS theory includes the Bayesian theory for statistically independent random variables as a special case, we have restricted our attention to independent random data generated using Bayesian-based techniques. This approach allowed us to generate the large number of data samples required to specify the desired probability mass functions. In addition, it enabled us to verify that the approach was suitable for a variety of underlying probability distributions. As noted in the literature (*1*), the number of unknown DS parameters increases rapidly with the number of measurement sensors and detection hypotheses. Hence, for this investigation, we restricted our attention to fusion of detection algorithm outputs from two sensors, thereby keeping the number of unknown parameters manageable. The techniques developed here, however, are readily extensible to a larger sensor suite and a larger number of target classes, if the increased computational requirements are tolerable.

The DS mass function is similar to the Bayesian probability density function (pdf). It assigns a number to a measurement (which could include a detection algorithm output), expressing the likelihood that the measurement indicates a specific object, such as clutter or target. In fact, we calculated the mass DS mass functions from the histograms for data samples from the "target" and "clutter" hypotheses. Calculating the mass assigned to the "don't know" hypothesis, however, is not so straightforward, and we spent a great deal of effort surmounting this problem.

We began our development for this project by considering the DS joint probability mass function (pmf) for measurements from two independent sensors. The joint pmf is the DS analog of the Bayesian joint pdf, and it can be expressed in terms of the mass functions for the individual sensors using the DS combination rule:

$$m_{12}(\mathbf{X}, \mathrm{H}_i) = \begin{cases} m_{12}(\phi) = 0 \\ = \dfrac{m_1(x_1, \mathrm{H}_0)m_2(x_2, \mathrm{H}_0) + m_1(x_1, \mathrm{H}_0)m_2(x_2, \mathrm{H}_0 \cup \mathrm{H}_1) + m_1(x_1, \mathrm{H}_0 \cup \mathrm{H}_1)m_2(x_2, \mathrm{H}_0)}{K}, i = 0, K \neq 0 \\ = \dfrac{m_1(x_1, \mathrm{H}_1)m_2(x_2, \mathrm{H}_1) + m_1(x_1, \mathrm{H}_1)m_2(x_2, \mathrm{H}_0 \cup \mathrm{H}_1) + m_1(x_1, \mathrm{H}_0 \cup \mathrm{H}_1)m_2(x_2, \mathrm{H}_1)}{K}, i = 1, K \neq 0 \\ = \dfrac{m_1(x_1, \mathrm{H}_0 \cup \mathrm{H}_1)m_2(x_2, \mathrm{H}_0 \cup \mathrm{H}_1)}{K}, i = 2, K \neq 0 \end{cases}, \quad (1)$$

where $m_{12}(\mathrm{X},\mathrm{H}_i)$ denotes the joint pmf for sensor 1 and sensor 2; H_i denotes hypothesis i (either "target," "clutter," or "uncertain"); X denotes the vector of measurement statistics (features) from sensor 1 and sensor 2; and $\mathrm{H}_0 \cup \mathrm{H}_1$ denotes the region of feature space assigned probability mass corresponding to "uncertain" or "don't know." K is a normalizing factor defined by

$$\begin{aligned} K = {} & m_1(x_1, \mathrm{H}_0)m_2(x_2, \mathrm{H}_0) + m_1(x_1, \mathrm{H}_0)m_2(x_2, \mathrm{H}_0 \cup \mathrm{H}_1) + m_1(x_1, \mathrm{H}_0 \cup \mathrm{H}_1)m_2(x_2, \mathrm{H}_0) + \\ & m_1(x_1, \mathrm{H}_1)m_2(x_2, \mathrm{H}_1) + m_1(x_1, \mathrm{H}_1)m_2(x_2, \mathrm{H}_0 \cup \mathrm{H}_1) + m_1(x_1, \mathrm{H}_0 \cup \mathrm{H}_1)m_2(x_2, \mathrm{H}_1) + \\ & m_1(x_1, \mathrm{H}_0 \cup \mathrm{H}_1)m_2(x_2, \mathrm{H}_0 \cup \mathrm{H}_1) \end{aligned} . \quad (2)$$

Our goal is to define a metric and a procedure that enable us to optimally assign probability mass to $\mathrm{H}_0 \cup \mathrm{H}_1$. One such approach used in the past measured the amount of uncertainty remaining after data fusion, declaring the pmf mass assignments to be good if this residual uncertainty is reduced (*2*), while another approach combined a maximum-likelihood parameter-estimation approach with DS belief functions (*3*). In what follows, we eschew the statistical quantities, such as plausibility and belief, which are often employed by practitioners of the DS theory (*4*). Instead, we present a novel approach that combines elements of DS-theory with classical detection-theoretic techniques commonly used within the automatic target recognition (ATR) community. That is, we modified the DS joint mass function to improve the performance of a test based on the classical likelihood-ratio test (LRT), and we quantified this performance via receiver operator characteristic (ROC) curves. It is this combination of the classical LRT, a ROC performance metric, and the DS probabilistic formulation that provides the novelty to our approach.

In the past, researchers have used the area under the ROC curve to compare the performance of two competing algorithms (*5*). We followed a similar approach and defined a fitness measure that expresses performance in terms of the area under the ROC between two pre-determined probabilities of false alarm (Pfas), defined as $r(i)$, where i denotes the iteration number. Hence, we optimized the DS parameters for specific operating points at the expense of potentially degrading performance at other, less desirable operating points.

We calculated the ratio-based test statistics according to

$$L(X) = \frac{m_{12}(X, H_1)}{m_{12}(X, H_0)} \ , \tag{3}$$

where $m_{12}(X, H_i)$ is the joint pmf described in equation 1, H_1 denotes the "target" hypothesis, and H_0 denotes the "clutter" or "non-target" hypothesis. Our likelihood ratio test then becomes, simply

$$m_{12}(X, H_1) \underset{<}{>} k\, m_{12}(X, H_0) \ \begin{cases} Declare\ H_1 \\ Declare\ H_0 \end{cases} \ , \tag{4}$$

where k is a pre-determined threshold selected to fix a specific operational Pfa or probability of detection (Pd). Note that for the case of zero uncertainty $m_{12}(X, H_0 \cup H_1) = 0$, and the likelihood ratio test reduces to the classical Bayesian likelihood ratio test.

We recognize immediately from equations 4 and 1 that we can adjust constituent parameters contributing to $m_{12}(X, H_i)$ and alter the test statistic at selected operating Pfas. Values of $m_i(x_i, H_j)$ for i=1,2 and j=0,1 are dictated by the underlying marginal pdfs, and we calculate them using standard histogram-based techniques. Values of $m_i(x_i, H_0 \cup H_1)$, however, provide us with the degrees of freedom necessary to improve the performance of the likelihood ratio test. These probability "masses" comprise the parameters that we optimize via dynamic training algorithms.

Many of the algorithms for determining optimal parameters rely upon recursive techniques based on gradient descent, and they typically attempt to minimize a differentiable cost function, such as mean-squared error. Since we use the area under the ROC curve as our fitness measure, our metric calculation is nonlinear and unsuitable for a recursive, gradient-based technique. As a result, we have identified and implemented an appropriate nonlinear optimization procedure that is referred to in the literature as the "particle swarm" algorithm (*6*). As its name suggests, the particle swarm algorithm mimics the dynamics of swarming insects. Its initial step involves the creation of N swarm members, each located at some point in the multi-dimensional parameter space. In our application, this multi-dimensional space consists of the DS mass functions for the "don't know" hypothesis, $m_i(x_i, H_0 \cup H_1)$, and examples of initial seeds for these weighting functions are shown in figure 1. We refer to the set of initial weighting functions as "basis functions," because they generate the final DS mass function. The fitness function—in our case, a version of the area under the ROC curve—is evaluated at each swarm member's location, and the maximum (global) fitness value together with its corresponding location are recorded. Each swarm member's location is then modified according to equation 5:

$$\boldsymbol{s}_j(i+1) = \boldsymbol{s}_j(i) + \mu\,(\boldsymbol{g}(i) - \boldsymbol{s}_j(i)), \tag{5}$$

where $\boldsymbol{s}_j(i)$ denotes the location of swarm member j within the parameter space at iteration i; $\boldsymbol{g}(i)$ denotes the "globally best" location found by all swarms up to time i; and $0 < \mu < 1$ is a constant governing the step size from $\boldsymbol{s}_j(i)$ in the direction of $\boldsymbol{g}(i)$. As the algorithm progresses, the members of the swarm wander through the parameter space in the direction of the ever-changing "globally best" location. Whenever a new maximum is found, both the new maximum and the new parameter value are recorded, and the corresponding swarm member is initiated to another randomly selected basis function. The algorithm terminates either after a fixed point in the parameter space has been found or a designated number of iterations have been completed. After the algorithm has converged, we are left with a final version of the mass function as illustrated in figure 2.

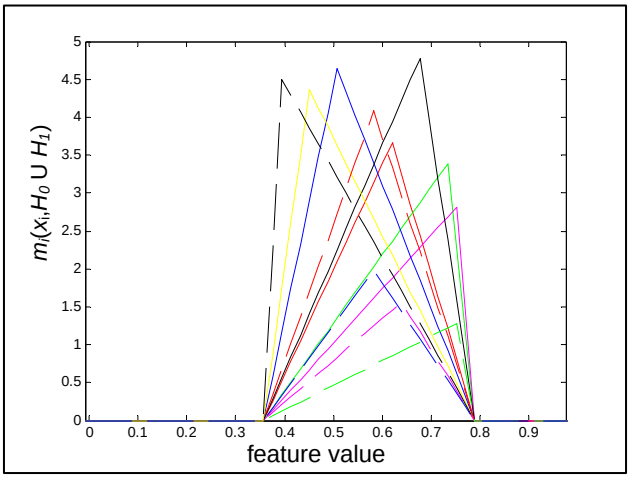


Figure 1. Example of members of a swarm at an initial iteration for $m_1(x_1, H_0 \cup H_1)$. The feature value is a normalized output from a notional ATD algorithm.

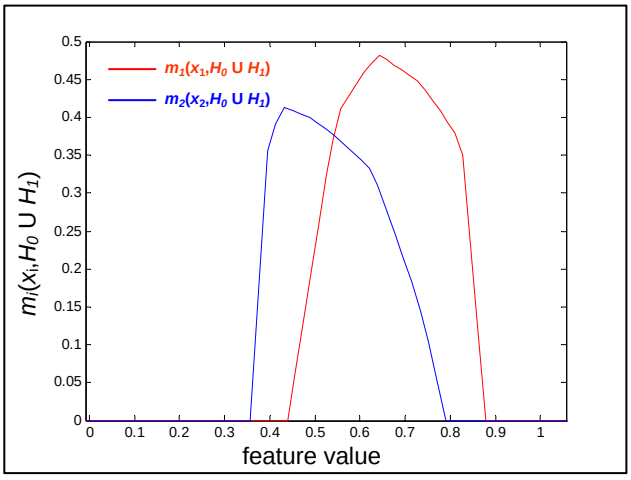


Figure 2. Example of final mass functions after the particle swarm algorithm has converged; here, i =1,2.

Both the number and initial disposition of swarm members play a critical role, since they determine how the search will proceed. We specify 300 randomly selected initial locations in the parameter space so as to bracket anticipated values of the unknown mass functions, and we specify a scalar value of μ that produces a step size judged to be adequate but not too large (on the order of $\|\boldsymbol{g}(i) - \boldsymbol{s}_j(i)\|/10$). The 300 functions comprise different triangular waveforms similar to those depicted in figure 1, and we also vary the magnitude of mass functions from sensor 1 relative to those from sensor 2. That is, we include three separate cases: one in which the mass values for sensor 1 are smaller than the corresponding mass values for sensor 2, a second in which the mass values are about the same size, and a third in which the mass values for sensor 1 are larger than the corresponding mass values for sensor 2. Once again, this is done in an attempt to bracket the region of the parameter space containing the solution. The entire particle swarm procedure is outlined in the block diagram in figure 3.

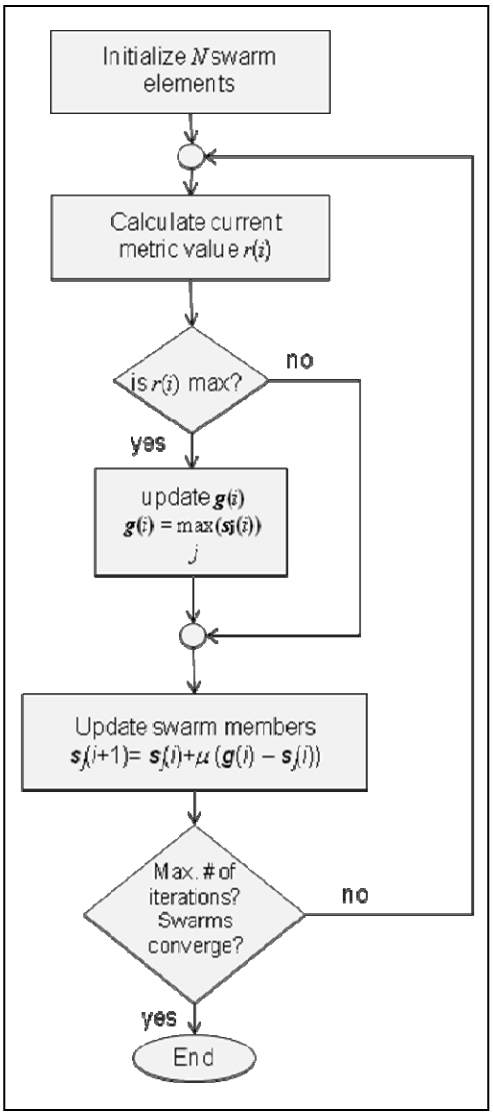


Figure 3. Block diagram of a particle swarm.

To evaluate our proposed technique, we simulated multiple sensing scenarios, varying the means, standard deviations, and correlations of the two-dimensional Gaussian-distributed feature vectors representing measurements from two independent sensors. The 10^6 samples representing clutter measurements from one of the sensors were independent and identically distributed (iid), and they may or may not have been correlated with the 10^6 samples representing clutter measurements from the other sensor. The 10^5 samples representing target measurements from one of the sensors were iid and were independent of the clutter samples. For some of the scenarios they were, however, correlated with the 10^5 samples representing measurements from the second sensor. Clutter plots of simulated measurements for representative scenarios are shown in figure 4.

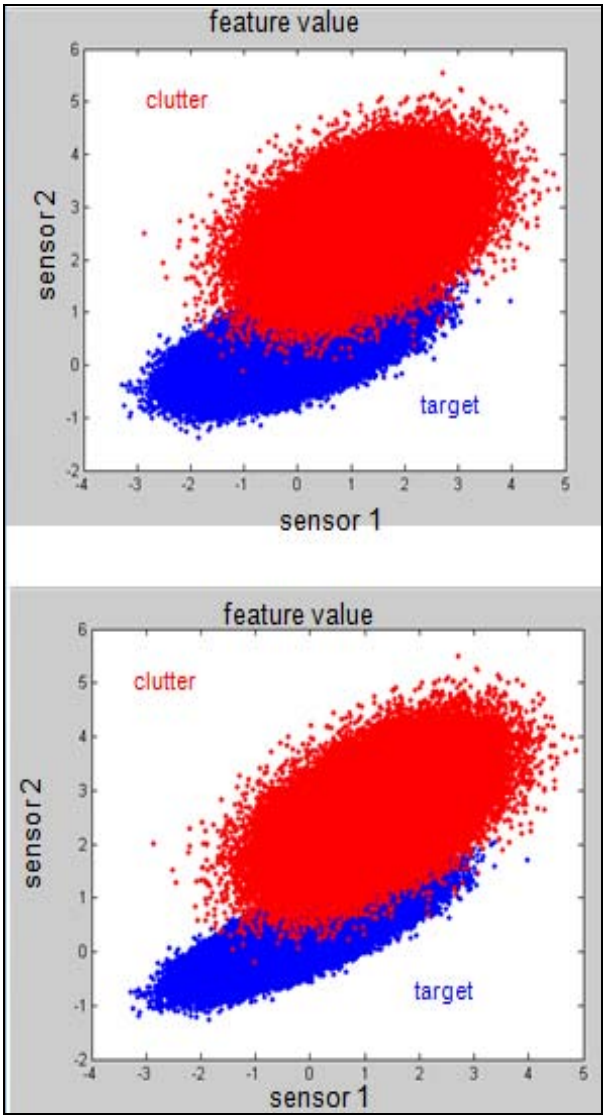


Figure 4. Scatter plots of sample scatter plots of sample clutter (red) and target (blue) data. The values for sensor 1 and sensor 2 represent outputs of notional ATD algorithm.

We calculated the mass function for a particular sensor from the histogram of relevant data samples in figure 4, and examples of these mass functions are shown in figure 5. Notice that the amount of correlation varies slightly between the two data sets. We defined a region of support for basis functions, such as those shown in figure 1, by examining the sensor measurements for which target and clutter samples have similar mass values in the plots of figure 5. For our investigations, we assigned the non-zero mass for $m_i(x_i, H_0 \cup H_1)$ (the uncertainty region) to sensor measurements with corresponding mass assignments in the interval [0.2,0.8]—a selection based on results of a brief parametric study conducted using one of the preliminary data sets. Note that we fixed the parameter k in equation 4 at 1.0 in order expedite the training process, but in the most general formulation it too would be determined by the training procedure.

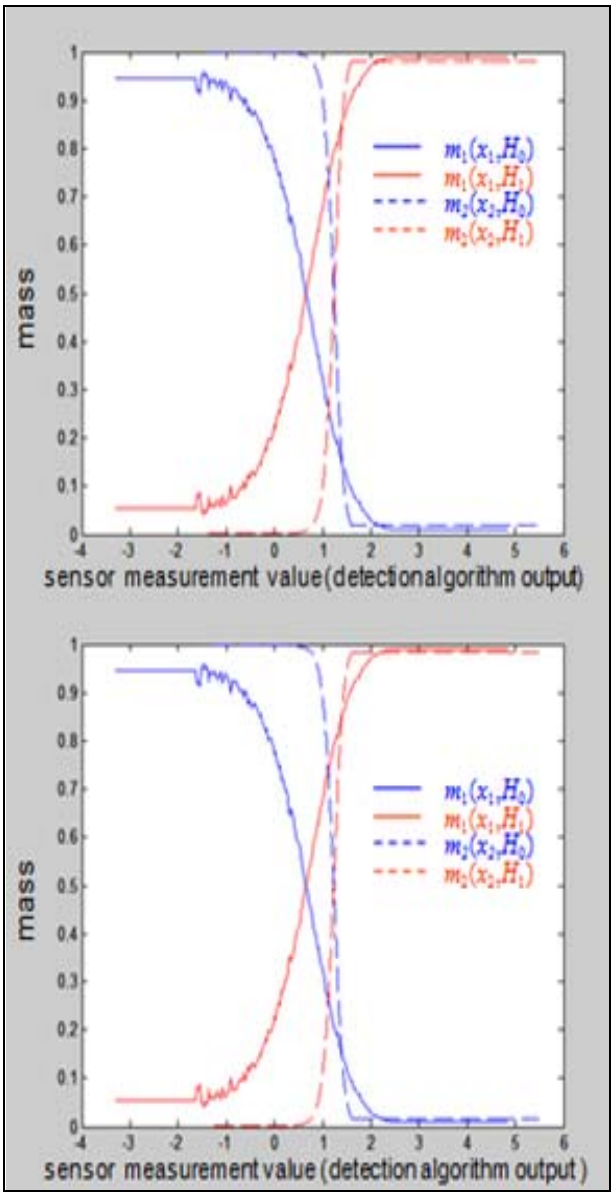
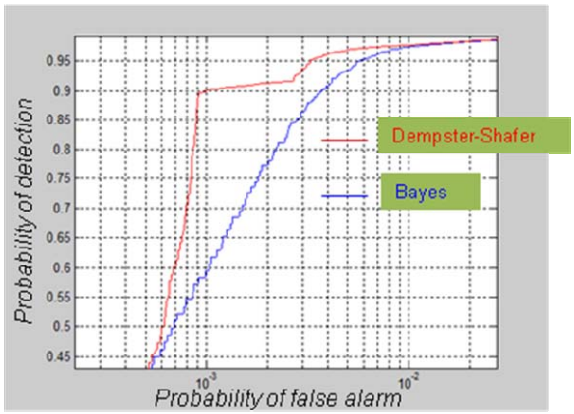


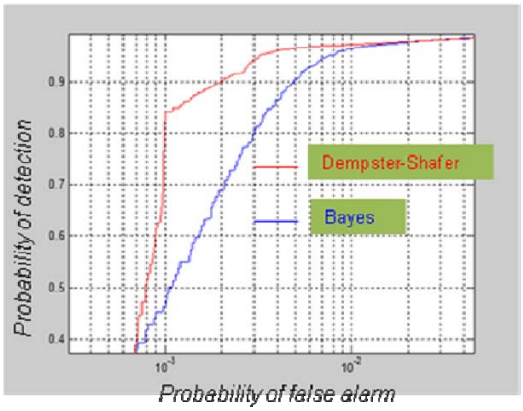
Figure 5. Mass functions corresponding to data in figure 4.

3. Results

We evaluated detection algorithm performance using data distributed as illustrated in figure 4. Our evaluation metric was the ROC curve, and we restricted our attention to the specific Pfa interval defined as part of the training procedure. Since our Pfas of interest were low, we used 4×10^6 clutter samples to ensure that we have a significant number of samples available for estimating these probabilities. Results are included in figure 6, along with the underlying detection algorithm and data generation parameters. Figure 7 shows the resulting uncertainty functions together with the probability mass functions for sensor 1 and sensor 2. We do not include the analyses for uncorrelated data, since these results did not indicate a significant difference between the Bayesian and the DS approaches. For uncorrelated data, the particle swarm converged to uncertainty functions that were zero or nearly zero everywhere.

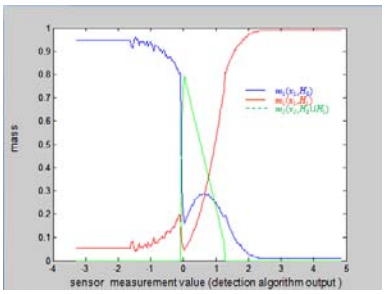


(a) Parameter set 1. σ_1=0.5, σ_2=0.13, σ_{12}=0.15 for clutter σ_{12}=0.2 σ_1=0.75, σ_2=0.42, σ_{12}=0.23 for target clutter mean=[0.0, 0.38], target mean = [1.25,2.62]

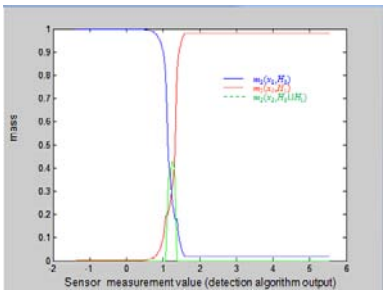


(b) Parameter set 2. σ_1=0.5, σ_2=0.13, σ_1=0.75, σ_2=0.42, σ_{12}=0.3 clutter mean=[0.0, 0.38], target mean = [1.25,2.62]

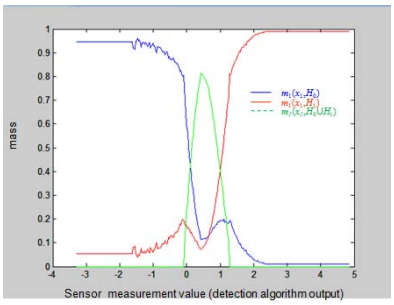
Figure 6. ROC curves generated for two different sets of algorithm and data parameter values; the data sets were used to generate plots in figures 4 and 5, and all data follow normal distribution.



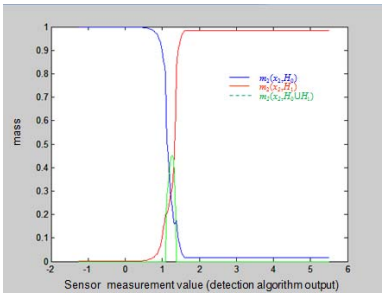
(a) DS mass functions for sensor 1, Parameter set 1.



(b) DS mass functions for sensor 2, Parameter set 1.



(c) DS mass functions for sensor 1, Parameter set 2.



(d) DS mass functions for sensor 2, Parameter set 2.

Figure 7. DS mass functions (including uncertainty function, in green) found by the particle swarm for parameter sets shown in figure 6.

4. Conclusions

The simulation results indicate that it is possible to enhance performance of the ratio-based detection algorithm over a designated Pfa interval; although this gain could result in decreased performance over a range of Pfas that are not of interest. The expected improvements in performance appear to be insignificant when the assumption of independence between the two sensors is satisfied, but this is not entirely surprising, since the classic LRT was derived under this assumption of independence. When correlation between the sensor measurements is introduced, however, the improvement obtained via DS can be significant. This improvement is due to the additional degrees of freedom available from the inclusion of the "don't know" category (for individual sensors) into the calculation of the DS joint mass function. While data-intensive, the DS method provides the possibility of enhancing detection algorithm performance at critical values of Pfa, and our procedure, combining concepts from DS and classical detection theory, exploits this additional flexibility. With an adequate amount of training data, the DS-based approach could become an attractive option.

5. References

1. Yager, R.; Fedrizzi, M.; Kacprzyk, J., eds. *Advances in the Dempster-Shafer Theory of Evidence*; John Wiley & Sons, Inc.: New York, 1994.

2. Zhu, Y. M.; Bentabet, L.; Dupuis, O.; Kaftandjian, V.; Rombaut, M. Automatic Determination of Mass Functions in Dempster-Shafer Theory Using Fuzzy C-means and Spatial Neighborhood Information for Image Segmentation. *Opt. Eng.* **April 2002**, *41* (4) 760–770, SPIE.

3. Zribi, M.; Benjelloun, M. Parametric Estimation of Dempster-Shafer Belief Functions, *Proceedings of the Sixth International Conference of Information Fusion*, Vol. 1, 2003.

4. Wu, H.; Siegel, M.; Stiefelhagen, R.; Yang, J. Sensor Fusion Using Dempster-Shafer Theory, *IEEE Instrumentation and Measurement Technology Conference*, May 21–23, 2002, 7–2.

5. Wilson, J. N. Use of Rank-based Decision Level Fusion in Landmine Discrimination. *Proceedings of SPIE,* Vol. 6953, 2008, 69531E1–69531E9.

6. Kennedya, J.; Eberhart, R. Particle Swarm Optimization. *Proceedings of the IEEE International Conference on Neural Networks*, Vol. 4, Nov 27–Dec 1, 1995, 1942–1948.

List of Symbols, Abbreviations, and Acronyms

ATD	automatic target detection
ATR	automatic target recognition
DS	Dempster-Shafer
iid	independent and identically distributed
LRT	likelihood-ratio test
Pd	probability of detection
pdf	probability density function
Pfa	probability of false alarm
pmf	probability mass function
ROC	receiver operator characteristic

Copies	Organization
1 (PDF only)	DEFENSE TECHNICAL INFORMATION CTR DTIC OCA 8725 JOHN J KINGMAN RD STE 0944 FORT BELVOIR VA 22060-6218
1 CD	DIRECTOR US ARMY RESEARCH LAB IMNE ALC HR 2800 POWDER MILL RD ADELPHI MD 20783-1197
1 CD	DIRECTOR US ARMY RESEARCH LAB AMSRD ARL CI OK TL 2800 POWDER MILL RD ADELPHI MD 20783-1197
1 CD	DIRECTOR US ARMY RESEARCH LAB AMSRD ARL CI OK PE 2800 POWDER MILL RD ADELPHI MD 20783-1197
3 HC	US ARMY RESEARCH LAB AMSRD ARL SE RU K RANNEY 2800 POWDER MILL RD ADELPHI MD 20783-1197
1 HC	US ARMY RESEARCH LAB AMSRD ARL SE RU K KAPPRA 2800 POWDER MILL RD ADELPHI MD 20783-1197
1 HC	US ARMY RESEARCH LAB AMSRD ARL SE RU N NASRABADI 2800 POWDER MILL RD ADELPHI MD 20783-1197

ABERDEEN PROVING GROUND

1 CD	DIR USARL AMSRD ARL CI OK TP (BLDG 4600)

TOTAL: 10 (5 HCs, 4 CDs, 1 PDF)